DOI:
10.1177/0278364919884623

Document Version
Peer reviewed version

Quantifying Teaching Behaviour in Robot Learning from Demonstration

International Journal of Robotics Research
XX(X):1–17

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Aran Sena[1] and Matthew Howard[1]

Abstract

Learning from demonstration allows for rapid deployment of robot manipulators to a great many tasks, by relying on a person showing the robot what to do rather than programming it. While this approach provides many opportunities, measuring, evaluating and improving the person's teaching ability has remained largely unexplored in robot manipulation research. To this end, a model for learning from demonstration is presented here which incorporates the teacher's understanding of, and influence on, the learner. The proposed model is used to clarify the teacher's objectives during learning from demonstration, providing new views on how teaching failures and efficiency can be defined. The benefit of this approach is shown in two experiments ($n = 30$ and $n = 36$, respectively), which highlight the difficulty teachers have in providing effective demonstrations, and show how $\sim 169 - 180\%$ improvement in teaching efficiency can be achieved through evaluation and feedback shaped by the proposed framework, relative to unguided teaching.

Keywords

Learning from Demonstration, Machine Teaching, Human Robot Interaction

1 Introduction

The rise of collaborative robots—robots designed to work safely in close proximity with people—is making it more common for people interacting with robots to have little or no technical background. These systems reduce the requirement for conventional programming skills for robot operation through the use of simplified programming interfaces, making deployment faster and more flexible. Of particular interest for deploying such systems is Learning from Demonstration (LfD) as a way to enable *novice* users (*i.e.,* people with no relevant technical background) to *teach* robots to perform useful work, similarly to how they would train an ordinary co-worker (Billard et al., 2008; Argall et al., 2009; Calinon and Lee, 2017).

Key to making LfD practical in the real world, is ensuring efficient skill-learning by the robot. Efficient skill-learning in this context means the robot is able to not just learn to perform the demonstrated task *as it was shown by the person*, but to *generalise beyond the specific demonstrations provided*, to new circumstances. This ability helps the robot to adapt to uncertainty during operation and reduces the teaching effort for the person doing the teaching.

Deploying systems that can learn generalised tasks from human teachers presents a number of challenges for both the learner and the teacher. When a system incorporates human input into its learning process, the system performance strongly depends on the quality of the data provided. As noted by Argall et al. (2009), key issues that can arise as a result of poor teaching in LfD include (i) *undemonstrated states*, (ii) *ambiguous demonstrations*, and (iii) *unsuccessful demonstrations*, all of which might confuse the robot learner, resulting in poor performance during skill realisation. Poor teaching is especially prevalent with novice teachers and

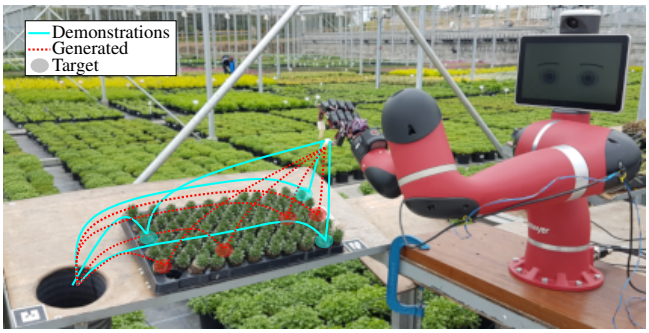


Figure 1. Learning from Demonstration can be used to learn a skill—here, disposal of unhealthy plants from a plant tray—from a limited set of user-provided demonstrations. Key to this is the ability to generalise from the demonstrated trajectories (cyan) to determine the appropriate trajectory to pick plants from previously unseen locations (red).

can be attributed to teachers lacking a shared mental model with the robot, or a lack of understanding of *how* and *what* the robot learns during the teaching process (Cakmak and Thomaz, 2014; Hellström and Bensch, 2018).

The *role of the teacher* in LfD systems is an under-explored area of research, with a lack of formal methods for *measuring, evaluating and modifying teaching behaviour*. To address this, a new theoretical framework for the assessment and guidance of LfD is presented here. Built

[1]Department of Engineering, King's College London, UK

Corresponding author:
Aran Sena, King's College London Strand, London, WC2R 2LS, UK.
Email: aran.sena@kcl.ac.uk

(a) (b)

Figure 2. "Low-batch" manual tasks found in the ornamental horticulture industry. (a) Quality assessment automation is often hampered due to companies growing a large variety of products, and (b) packaging often requires manual labour due to varying retail packaging requirements that can change seasonally, thus creating a need for more flexible automation systems.

on prior work in machine teaching, active learning, and LfD, the proposed framework enables a formal analysis of human-robot teaching during LfD to help people more effectively teach robots new skills. It facilitates (i) the definition of metrics to assess users' quality of teaching, (ii) the quantitative measurement of teaching failures, such as undemonstrated states and ambiguous demonstrations, and (iii) aids the development of tools (visualisations, procedures, *etc.*) that can provide a clearer view of how to guide the teaching process.

The effectiveness of this framework as a means of helping novices during LfD is shown in two experiments with novice users, extending the work presented in (Sena et al., 2018). It is shown in these validation experiments that the evaluation and feedback tools derived from the proposed framework allow for improving teaching performance, in the order of a 180% improvement relative to unguided conditions. This highlights the benefit of measuring, evaluating, and modifying *human teaching practice* during LfD, and provides insight into how LfD could be more widely used as a practical tool for deploying robotic automation.

2 Background and Related Work

Presented here is background on the field of LfD, and the current gap in considering the role of the teacher. Insights from the related domain of algorithmic machine teaching are then considered, to point to new directions for LfD modelling.

Learning from demonstration in robotics is often described as a form of supervised learning, where a teacher provides examples of a target task that the robot uses to learn a control policy (Billard et al., 2008; Argall et al., 2009; Billing and Hellström, 2010). LfD is a useful approach for robot task learning as it reduces the complexity of search spaces for learning real-world tasks, and reduces the amount of tedious programming required which helps *novice users* use robot systems (Billard et al., 2008; Chernova and Thomaz, 2014).

There is a large body of research available on the policy learning aspect of LfD, including symbolic reasoning methods (Billard et al., 2008; Ahmadzadeh et al., 2015), reinforcement learning based methods (Schaal, 1996; Ng and Russell, 2000; Abbeel and Ng, 2004; Kormushev et al., 2013), dynamical system modelling methods (Schaal, 2006; Pervez and Lee, 2017), probabilistic methods (Asfour et al., 2006; Calinon and Billard, 2007b; Cederborg et al., 2010; Calinon, 2015; Maeda et al., 2017; Huang et al., 2018), particle based approaches (Groth and Henrich, 2014; Orendt et al., 2016), and geometric based methods (Ahmadzadeh and Chernova, 2018). There is, however, less research found in robot LfD on the teaching provided by people to robots, particularly for learning grasping tasks.

While expert-level knowledge may not be required with LfD, with many possible policy learning methods, there will be many different requirements for the data provided to the learning system. The optimal teaching strategy a user should use may therefore be difficult to identify for novice users.

2.1 Teaching Feedback and Evaluation

Feedback from robot learners to human teachers in LfD is often considered in pragmatic terms, with solutions determined based on the task at hand, or simply assuming the teacher will be able to successfully interpret the learner's actions and adjust their teaching behaviour accordingly. Incremental methods in LfD are viewed as a way of gradually learning a skill, and can be adapted to help improve and/or overcome a users physical skill deficiencies (Calinon and Billard, 2007a; Hoyos et al., 2016; Tykal et al., 2016). However, even though the process is iterative, there has been little research on how novice users interpret and adapt to the learner over consecutive teaching steps.

Calinon and Billard (2007a) and Weiss et al. (2009) both describe an incremental learning system where the active role that teachers can play during the learning process is considered. In deciding where to provide a new demonstration, the authors describe the teacher observing the robot attempt the taught skill in new locations to determine how to provide the next demonstration. This overlooks the question of *how* the attempts should be selected to help inform the teacher, whether the teacher will be able to effectively test the learner's knowledge, and whether the teacher will be able to correctly interpret the attempts.

Effective feedback of what a robot has learned to do can allow a teacher to provide more effective instruction to the learner, without the need for understanding how learning is taking place. This is shown in Nicolescu and Mataric (2003), where teachers observe robots executing learned plans. Similarly, Argall et al. (2007) describe a system for allowing a teacher to provide specific feedback on trajectories generated by a learner after teaching. This highlights the benefit of informed teaching from a human teacher, made possible through effective feedback of what the robot has learned to do.

Toris et al. (2012) presents a user study comparing three LfD policy learning methods on a sweeping task. A common feedback point raised by participants was the need for a better understanding of what the robot is "thinking". While direct comparison of learning methods for usability is valuable, it is likely that specific methods will be of benefit to specific types of tasks.

An additional benefit to improving feedback in LfD systems, for the benefit of helping the teacher to understand the learner's capabilities, is this helps to facilitate trust between the human teacher and robot learner (Yang et al., 2017; Lewis et al., 2018), though this is not directly explored in this work.

Other approaches to achieving skill learning and generalisation with minimal teaching effort have considered *active learning* based approaches, where the learner decides when a new demonstration is required. This can be achieved by either randomly sampling trajectories and asking the teacher if they are acceptable, or requesting new demonstration when entering regions of high uncertainty (Argall et al., 2009; Cakmak and Thomaz, 2011; Maeda et al., 2017). While this may be a complimentary approach to improving teaching, this does not supersede the benefit of a good teacher, as discussed in §2.2, and thus the need to understand how to measure and improve teaching remains important.

Similar to issues found in LfD feedback, evaluation of LfD systems has tended to focus on robot learners. Evaluation of teachers has more focused on their physical skill when executing the task, such as in Ureche et al. (2015) where metrics are proposed for determining quality of bimanual demonstrations, and Cho and Jo (2013) where an approach for identifying good teaching is determined by identifying whether provided demonstrations are consistent with previously provided demonstrations. While this could help a learner avoid taking demonstrations from a bad teacher, this does not directly evaluate the quality of teaching and so does not necessarily help the teacher improve.

The proposed framework therefore intends to emphasise the teacher's contribution to learning, and provide new approaches for designing feedback and evaluation tools in LfD systems.

2.2 Insights from Machine Teaching

The ability to analyse teaching behaviour is critical to improving LfD performance. It can be shown that an optimal teacher can theoretically provide the minimum number of samples required to teach a learner a task, called the *teaching dimension* (Goldman and Kearns, 1995; Balbach and Zeugmann, 2009; Khan et al., 2011; Cakmak and Thomaz, 2014; Zhu, 2015), by providing *non-i.i.d.* samples to the learner which exploit the task structure and learning method employed. In toy-problems, the learning method and task structure can be clearly defined for analysis, but when using LfD for learning real-world tasks from novice users this is rarely the case and so the analysis methods employed in machine teaching are not directly applicable.

While identifying theoretically optimal teachers may not be possible in general LfD tasks, it is shown that providing feedback and guidance to the teacher to help them become informed about the learning process can change participants' teaching strategy, resulting in improved teaching performance (Zang et al., 2010; Cakmak and Thomaz, 2014). It is this feedback mechanism for improving the teacher that the here proposed framework aims to enable, by providing a structure with which LfD problems can be analysed, and with which support tools for teaching can be designed.

2.3 Modelling Learning from Demonstration

There have been a number of efforts in the wider human robot interaction (HRI) domain to model human behaviour and cognition when interacting with robots. These have typically focused on collaborative human-robot teams that work together to complete a task, rather than the problem faced in LfD of teaching new skills (Goodrich and Schultz, 2007; Nikolaidis et al., 2017; Hiatt et al., 2017).

Looking more specifically at HRI involving uncertain, or noisy, communication, Hellström and Bensch (2018) present a more general framework based on message compression. Here they discuss the robot's ability to influence the person through communication, however, being a general framework for HRI, it serves as a complementary resource to the more focused LfD-specific framework presented here.

While there have been several works highlighting the benefits of considering the teacher as an active contributor to the learning process in LfD, as discussed, there is a lack of structure in considering their role. A significant previous effort to introduce a formal structure to LfD can be found in (Billing and Hellström, 2010), where the authors model the LfD process following a message compression scheme. While this helps to provide a framework around the learning process in LfD, there is no consideration for how interaction with the system itself affects the teacher's actions. By not considering the influence of the teacher, the current approaches of focusing on the learner neglects opportunities for improving task learning performance in LfD.

An extension from Billing and Hellström (2010) is considered in Cederborg and Oudeyer (2014), where the authors present a generalised framework which considers multiple modalities of teacher feedback to the robot, with the view of using ambiguous teaching cues such as gaze to infer the learner's goal. Though the teacher is assumed to be infallible, the learner is presented as a passive receiver of information from the teacher in this case, missing the opportunity to actively improve the teacher's understanding of the learning process beyond explicit training by an expert.

Given the lack of a distinct framework which captures the teacher's contribution to the LfD process explicitly, and the benefit of doing so as suggested in prior works, there is a clear need to develop an improved formalism. By considering the teacher explicitly, it would be possible to define teaching objectives and quality measures in a task-agnostic representation, and design tools that allow the robot to more effectively guide the teacher toward better demonstrations.

3 Learning and Teaching Framework

In this section, a new, formal framework for assessing the role of the teacher in robot LfD is defined, along with the means by which it can be used to measure, evaluate, and

modify teacher behaviour. The proposed framework builds on the prior work of Billing and Hellström (2010), but diverges in that it places much greater emphasis on the role of the teacher.

3.1 Framework Definition

The proposed framework describes the LfD process as a set of mappings between an information history space, $\mathfrak{I}_h$, a policy space, $\mathbf{\Pi}$ and a teacher belief space $\mathfrak{M}$.

$\mathfrak{I}_h$ represents the space of all possible *event histories*, where each *element* is a tuple containing initial conditions, observations, and action sequences. In Figure 1 each of the red and cyan trajectories would represent one element in $\mathfrak{I}_h$, but also *incorrect* trajectories for this task, and trajectories for *other tasks* would also be represented by elements of $\mathfrak{I}_h$. That is, $\mathfrak{I}_h$ is defined for the possible actions of the robot, rather than exclusively being defined for one task.

$\mathbf{\Pi}$ represents the space of *all possible policies*, learnable by a selected learning function λ.

$\mathfrak{M}$ captures the *teacher's belief* about the extent to which the robot has learnt the task, and the demonstrations necessary to improve its performance. An overview of the framework is presented in Figure 3, and the following highlights some of its salient features.

3.1.1 Task Teaching — The space representing a behaviour, or task, that the user would like to teach the robot is denoted $\mathfrak{B}$, the *task space* (see Figure 3).

The elements of $\mathfrak{B}$ represent all possible ways of performing the task successfully, *i.e.,* $\mathfrak{B}$ forms a subset of $\mathfrak{I}_h$, the space of all possible things that the robot could do. A important note here is that membership of $\mathfrak{B}$ only considers whether trajectories are *feasible*, rather than making an assessment of *optimality*.

For example, in the plant grading task in Figure 1, $\mathfrak{B}$ might represent the set of all possible trajectories for picking and disposing of plants from any location in a tray. The learner does not have access directly to $\mathfrak{B}$, instead relying on a set of M demonstrations provided by the teacher, $\mathfrak{b}$. In the plant grading task example, $\mathfrak{b}$ consists of the set of demonstrated trajectories for picking plants from specific locations (cyan trajectories in Figure 1). The goal of improving teaching efficiency is therefore conceptually captured in this framework as that of *providing the best possible* $\mathfrak{b}$ *to enable the robot to learn the target behaviour* $\mathfrak{B}$. A formal metric for this is discussed below (refer to §3.2).

3.1.2 Task Learning — The framework assumes that the robot is equipped with learning capability, such that it can use demonstrations to form a task model. Specifically, there exists a learning function, λ that uses $\mathfrak{b}$ to derive a controller $\pi = \lambda(\mathfrak{b}) \in \mathbf{\Pi}$.

In real-world tasks, it is typically the case that* $|\mathfrak{b}| \ll |\mathfrak{B}|$, therefore the learning system must learn a model which can *generalise* from the subset $\mathfrak{b}$ provided. It is also typical that learning occurs as an *iterative operation*, that is, the learnt model is sequentially updated whenever new demonstrations are provided. The framework captures this by assuming that λ makes use of all past demonstrations provided, *i.e.,* $\pi = \lambda(\mathfrak{b}_m) \in \mathbf{\Pi}$, where $\mathfrak{b}_m$ contains the m demonstrations made available.

3.1.3 Task Realisation — Once a controller has been learnt, it can be executed by the learner to (try to) perform the task. Execution of a task using a learnt model is denoted as a mapping from $\mathbf{\Pi}$ to $\mathfrak{I}_h$ through a "realisation function" Λ, where each execution results in a *task realisation*, $\mathfrak{r} = \Lambda(\pi) \in \mathfrak{I}_h$. Here, $\mathfrak{r}$ represents the set of task realisations which are *actually observed* by the teacher.

In the plant grading example task realisations would represent the robot using its learnt model to generate picking trajectories (shown in red in Figure 1). The realisation space, $\mathfrak{R}$, represents the set of all possible realisations of the learnt model under different task conditions.

It is desired that the final learnt model would result in $\mathfrak{B} \subseteq \mathfrak{R}$, which is the case that the target task has been learnt completely. Billing and Hellström (2010) posit that the *learning objective* be described as the minimisation of $\mathfrak{B} \setminus \mathfrak{R}$ and $\mathfrak{R} \setminus \mathfrak{B}$ (*i.e.,* equality), meaning that the robot would reproduce the skill completely, and exclusively to, the entire space of the target behaviour. However, in practice, this objective is both (i) difficult to achieve (since in many cases, the spaces $\mathfrak{R}$ and $\mathfrak{B}$ are too large to fully observe, refer to §3.1.3), and (ii) overly prescriptive of the robot's behaviour. Specifically, in terms of the latter, it is often the case that learner behaviour in conditions outside the desired task, *i.e.,* $\mathfrak{R} \setminus \mathfrak{B}$, can simply be ignored. For example, in the plant grading task in Figure 1, the robot might learn how to pick plants *in close proximity, but external, to the tray* as a consequence of learning to pick those in the tray edge positions (a case of $\mathfrak{R} \setminus \mathfrak{B} \neq \emptyset$). This, however, does not conflict with performing the target skill, nor increase teaching effort.

Note, however, it is typically the case that $|\mathfrak{r}| \ll |\mathfrak{R}|$. This means that $\mathfrak{R}$ is not readily observable in practice, hence measurement of learning performance may require approximations.

3.1.4 Generalisation — In order for the learner to learn the target task effectively and efficiently, it is required for the learner to *generalise* from the provided demonstrations. Under the presented framework, generalisation is defined as

$$(\mathfrak{R} \setminus \mathfrak{b}) \cap \mathfrak{B} \neq \emptyset. \tag{1}$$

In other words, there exist task reproductions that do not occur in the demonstrated data $\mathfrak{b}$, but do fall within the definition of successful task performance $\mathfrak{B}$. This is captured by the striped region in Figure 5.

3.1.5 Modelling the Teacher — To develop an intuition for teaching behaviour and, importantly, why the teacher can fail to provide adequate demonstrations, the framework introduces a third space modelling the *teacher's beliefs about the learning process*. This is represented by a *belief space*, $\mathfrak{M}$ (see Figure 3).

As the teacher has no direct way of knowing the learner's state, $\mathfrak{R}$, they must estimate it based on the available information, *i.e.,* the realisations that they observe, $\mathfrak{r}$. This information is combined with the teacher's unseen factors that influence human behaviour, $\mathcal{Q}$, to form an *interpretation* of the learner state $\tilde{\mathfrak{R}}$. Here, $\mathcal{Q}$ captures difficult to measure

*Throughout this paper, $|\mathfrak{a}|$ is used to denote the cardinality of the set $\mathfrak{a}$.

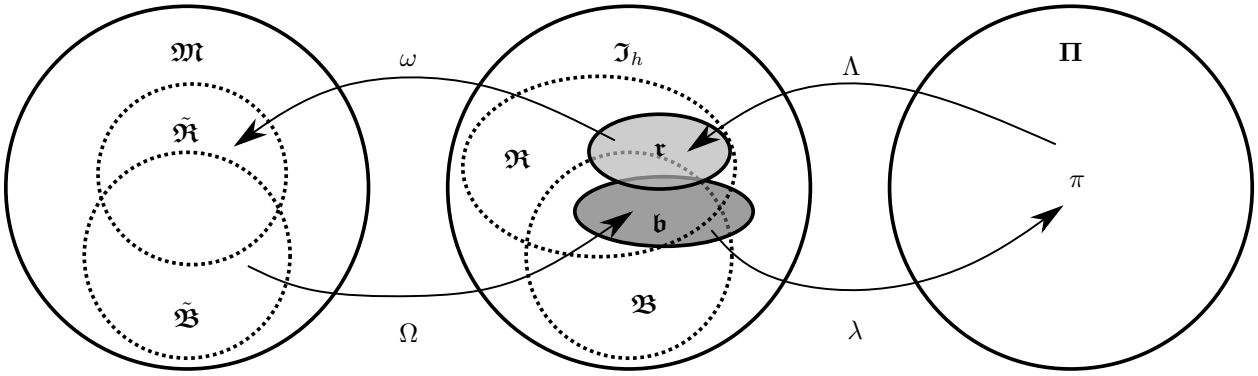


Figure 3. Learning from Demonstration model incorporating imperfect teaching. The space introduced on the left, $\mathfrak{M}$, represents the user's understanding of what the robot has learned so far, and what they would like the robot to learn. The shaded regions indicate parts of the process which are directly observable, *i.e.,* the actual demonstrations provided $\mathfrak{b}$ and the task realisations $\mathfrak{r}$.

factors such as mental state, or the person's prior expectation for how learning is taking place. Interpretation of $\tilde{\mathfrak{R}}$ is modelled as a mapping from $\mathfrak{I}_h$ to an estimated realisation space $\tilde{\mathfrak{R}} \subset \mathfrak{M}$ through an interpretation function

$$\tilde{\mathfrak{R}} = \omega(\mathfrak{r}, \mathcal{Q}). \tag{2}$$

In the context of iterative teaching, this estimated realisation space is then used to guide the teacher's next demonstration, in conjunction with their internal idea of what task the robot must learn $\tilde{\mathfrak{B}}$, and their human factors $\mathcal{Q}$. This is modelled as a mapping from $\mathfrak{M}$ back into $\mathfrak{I}_h$ through a demonstration function

$$\mathfrak{b}_{m+1} = \Omega(\tilde{\mathfrak{R}}, \tilde{\mathfrak{B}}, \mathfrak{b}_m, \mathcal{Q}). \tag{3}$$

It is expected that Ω is *not* a deterministic function, as $\mathcal{Q}$ will introduce variability in the teacher's behaviour; however generally it is expected that this will not significantly affect the performance evaluation of a given teacher, *i.e.,* it would still be possible to distinguish between good teachers and bad under typical conditions.

With the addition of $\mathfrak{M}$, it is possible to close the loop on the LfD process with a consideration for the teacher's internal belief processes and begin to reason about how teaching failures occur in this iterative cycle, see Figure 4.

3.1.6 Objective Task Approximations If the task space is very large or continuous, some form of approximation $\hat{\mathfrak{B}} \sim \mathfrak{B}$ and $\hat{\mathfrak{R}} \sim \mathfrak{R}$ may be required for measurement of teacher performance in practical situations. Note that the approximations $\hat{\mathfrak{B}}$ and $\hat{\mathfrak{R}}$ are distinct from the teacher's internal evaluation of the task and learner performance ($\tilde{\mathfrak{B}}$ and $\tilde{\mathfrak{R}}$, respectively) since they are *objective*, and therefore independent of the individual teacher's human factors (*i.e.,* are not influenced by $\mathcal{Q}$).

Such approximations may be formed through several approaches, from state-space reduction via discretisation and bounding, to statistical approximation with Monte Carlo methods. For example, in the plant grading task (shown in Figure 1), $\hat{\mathfrak{R}}$ can be defined as the trajectories generated given a pre-defined set of 'test plant' locations in the tray, enabling a measure of the learnt policy's performance to be made, without having to exhaustively test every possible plant location and every admissible approach trajectory for any given plant.

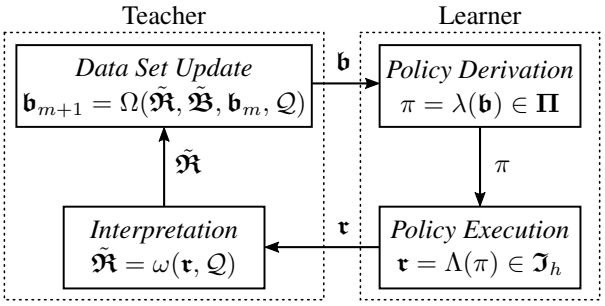


Figure 4. LfD pipeline, accounting for teacher influence. The teacher's interpretation of the robot's performance, $\tilde{\mathfrak{R}}$ is based on the task realisations that they actually observe, $\mathfrak{r}$, and hidden human factors, $\mathcal{Q}$. The way in which the teacher provides new demonstrations then depends on this interpretation, their interpretation of target behaviour $\tilde{\mathfrak{B}}$, and $\mathcal{Q}$.

The following sections describe measures of teaching performance and teaching failure modes in the general case of $\mathfrak{B}$ and $\mathfrak{R}$, however, these are equally applicable in the case that approximated spaces are used.

3.2 Evaluating Teaching by Demonstration

While evaluating the performance of the *learner* in LfD is well established, relatively little attention has been given to the performance of the teacher. The primary objective of the teacher is to ensure the learnt model allows the robot to execute the target skill. This primary objective is formally defined using the proposed feedback, along with further measures which are useful for evaluating teacher performance and identifying teaching failures.

3.2.1 Teaching Efficacy — The primary objective of teaching is to have the robot accurately learn the target skill, such that it is able to perform the desired task. From the perspective of *learning*, a necessary condition for this is that $\mathbf{\Pi}$ contains a policy which can produce trajectories sufficiently close to the target skill, and that λ applied to the demonstrations $\mathfrak{b}_m$ correctly identifies this policy.

Following discussion of what it means for the learner to learn in §3.1.3, the metric proposed here to evaluate *teaching efficacy* is

$$\varepsilon = \frac{|\mathfrak{R} \cap \mathfrak{B}|}{|\mathfrak{B}|}, \quad \varepsilon \in [0, 1]. \tag{4}$$

In other words, the goal is to achieve $\mathfrak{B} \subseteq \mathfrak{R}$, while ignoring what the policy has learned outside of the target space (*i.e.,* $\mathfrak{R} \setminus \mathfrak{B}$), normalised by the size of the task space $\mathfrak{B}$. This measure can be treated as an objective measure of the teacher's ability in enabling the learner to acquire the skill needed for the task.

The teacher's performance can then be monitored during interaction with the robot by considering the difference in efficacy with each additional demonstration.

3.2.2 Teaching Efficiency — Having defined the measure of teaching efficacy, it is then possible to consider the *teaching efficiency*. Efficiency in any given application is often context-dependent, however, typically, it is desirable to *minimise the total number of demonstrations required* since this is correlated with both the time spent teaching and the space needed to store data. There are other approaches to defining the efficiency, as explored in the related domain of *machine teaching* (Zhu, 2015), where alternatives to cardinality of the demonstration set as a measure of *teaching effort* are explored, such as penalising similarity of examples.

With this in mind, *teaching efficiency* is defined here simply as efficacy normalised by the number of demonstrations provided

$$\eta = \frac{\varepsilon}{|\mathfrak{b}|}, \quad \eta \in [0, 1]. \tag{5}$$

In other words, to be efficient, the teacher must achieve the maximum efficacy with the fewest possible demonstrations. Importantly, *ambiguous demonstrations*, *undemonstrated states*, and *incorrect demonstrations* will all result in reducing the teaching efficiency metric, as discussed below.

The definitions (4) and (5) provide a way to monitor a user's teaching performance. In the next section, commonly-encountered teaching failures are analysed in light of the new framework.

3.3 Understanding Teaching Failures

With the framework established, it is now possible to make formal definitions of failure modes in LfD teaching, and metrics for their quantitative evaluation. Specifically, the below examines three common teaching failures, namely, (i) *incorrect demonstrations*, (ii) *ambiguous demonstrations*, and (iii) *undemonstrated states*. Each of these can be attributed to poor teacher skill, Ω, affecting the quality of data provided to the learner, or poor user judgement, ω, affecting the accuracy of the teacher's estimation $\tilde{\mathfrak{R}}$ (see §3.1.5).

3.3.1 Incorrect Demonstrations — The most fundamental teaching failure is that of providing *incorrect*, or unsuccessful, demonstrations, *i.e.,* demonstrations that are *not* examples of the target skill. Formally, these can be defined as

$$\mathfrak{b} \setminus \mathfrak{B} \neq \emptyset \tag{6}$$

illustrated by the solid-shaded region in Figure 5.

The effect of incorrect demonstrations can be observed through a reduction in teaching efficacy, as the learner's performance will degrade

$$\varepsilon_m - \varepsilon_{m-1} \leq 0. \tag{7}$$

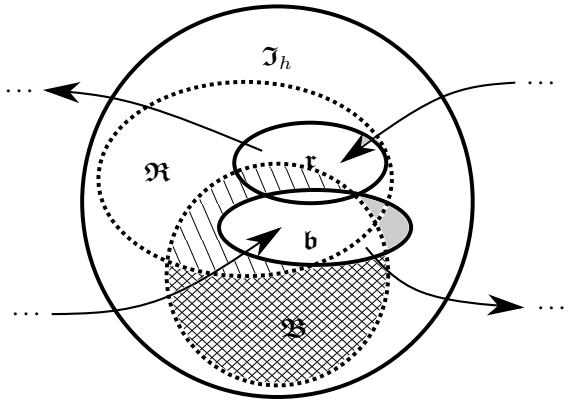


Figure 5. Focusing on the information history space, three subsets of interest can be identified. The striped region in the centre of the diagram represents parts of the task which have been correctly learned through *generalisation* from the demonstration set. The cross-hatched region then represents undemonstrated regions of the task. Finally the shaded region represents demonstrations that the teacher has performed incorrectly. See §3.2.

Incorrect demonstrations can occur for a number of reasons. For instance, the teacher may have poor skill in operating the robot, possibly indicating they need more time interacting with the robot to become accustomed to it. Alternatively, the teacher may make errors, such as forgetting which object they had decided to demonstrate for the robot, mid-demonstration. In both cases, incorrect demonstrations represent a failure in the demonstration function, Ω, as a result of human factors specific to the teacher, $\mathcal{Q}$, in the provision of demonstration data (see Figure 4). While it may be difficult to pinpoint a particular reason for an incorrect demonstration, (7) provides a measure for identifying an incorrect demonstration.

A note on incorrect demonstrations is that they will not always strictly result in a degradation of performance. Given a sufficiently sophisticated learner that can learn from incorrect demonstrations, *e.g.,* (Grollman, 2012), useful learning may still take place.

3.3.2 Undemonstrated States — As mentioned in §3.1.2, usually $|\mathfrak{b}| \ll |\mathfrak{B}|$ and it is not expected that a teacher could provide $\mathfrak{B}$ directly. It is therefore inevitable that the learner must generalise as much as possible from the demonstration set to effectively perform the desired task. If there are relevant states in which the robot cannot perform the task, after the teacher has provided all of the demonstrations they *think* are required for the robot to learn, these remaining states are referred to as undemonstrated states.

Undemonstrated states occur as there is a limit to the extent which generalisation is possible for a given $\mathfrak{b}$, and so the teacher must form good estimations of the robot's learned ability to identify when enough data has been provided, *i.e.,* ideally $\tilde{\mathfrak{R}} \sim \mathfrak{R} \sim \mathfrak{B}$.

The formalism defines undemonstrated states as the set

$$\mathfrak{B} \setminus (\mathfrak{R} \cup \mathfrak{b}) \neq \emptyset \tag{8}$$

illustrated by the cross-hatched region in Figure 5.

From the perspective of the teacher, failure to adequately demonstrate the task, such that the learner is able to

effectively perform the task in all expected conditions ($\mathfrak{B}$) is a sign that the teacher does not understand the learner's current abilities, *i.e.,* they have a poor estimate of $\tilde{\mathfrak{R}}$ where they are *overestimating* the learner's ability to perform the task. Asking the teacher to provide more demonstrations is not necessarily a solution to this failure, as practically $|\mathfrak{b}| \ll |\mathfrak{B}|$, plus the teacher may not be aware where the learner's performance is deficient. Pathways to improving this estimate may include improving the selection of task realisations $\mathfrak{r}$ to give more useful information to the teacher so they can improve $\tilde{\mathfrak{R}}$, or explicitly training the teacher on how the robot learns in order to give them more accurate estimates of the learner, *i.e.,* influencing $\mathcal{Q}$ to improve $\tilde{\mathfrak{R}}$ as described in (2).

3.3.3 Ambiguous Demonstrations — The third teaching failure that commonly occurs in LfD scenarios is the issue of *ambiguous demonstrations*. These are defined as demonstrations which offer little or no new information to the learner, such that performance on the target task is unchanged. This can happen, for example, when demonstrations provided by the teacher are very similar to those already been seen by the learner.

As ambiguous demonstrations result in little or no improvement in the learner's performance, and also do not significantly degrade it, ambiguous teaching can be identified by checking whether learner efficacy lies within an upper and lower bound

$$\delta_l \leq \varepsilon_m - \varepsilon_{m-1} \leq \delta_u. \tag{9}$$

Depending on the learning model being employed, if generalisation is being taught through varying demonstrations or if it is difficult for a learner to distinguish between similar examples, an ambiguous demonstration can be identified without updating and evaluating a learned model by evaluating demonstration *similarity* to previously provided demonstrations, within some threshold

$$s(b, \mathfrak{b}) \leq \delta_a, \tag{10}$$

where b is the new demonstration, $s(\cdot,\cdot)$ is some measure of similarity of b with respect to the existing demonstrations in $\mathfrak{b}$, and δ_a is an ambiguity threshold level for s. For instance, s could be the mean Euclidean distance difference between a demonstrated trajectory and each of those in the data set $\mathfrak{b}$, and δ_a set at some minimum threshold for this.
Similar to undemonstrated states, ambiguous demonstrations can be attributed to the teacher forming a poor estimate of the learner's ability, $\tilde{\mathfrak{R}}$, however, in this case, the teacher is *underestimating* the impact individual demonstrations are having on the learner's ability.

To summarise, the proposed framework improves upon the previous models of LfD by explicitly modelling what is observed by the human teacher, and how this might be interpreted by them. With this modification, definitions for measures of teacher performance and teaching failures naturally follow. With these performance measures in place, it is possible to design feedback tools with the specific purpose of influencing the teacher's belief space, $\mathfrak{M}$, which can be used to guide teachers toward more effective teaching practices. This is explored in the following two experiments, demonstrating the benefits of the proposed framework.

4 Evaluation

In this section, two experimental studies in LfD are presented to explore how the proposed framework can be used to measure, evaluate, and modify teacher behaviour to improve the overall system performance. Specifically, the experiments examine how it allows the design of effective feedback approaches that help the teacher understand the learner by improving interpretation, ω, to give the teacher better estimations of the learner's ability, $\tilde{\mathfrak{R}}$. By closing this loop between teacher and learner, it is hoped that the teacher will be able to provide higher quality demonstrations, $\mathfrak{b}$, during the data set update step, and thus increase the performance of the robot compared to an unguided teacher.

4.1 Experiment 1: Point-to-Point Reaching

In the first experiment, the teacher must teach a robot to navigate a maze from a start region to a goal (see Figure 6a). This task is chosen as it represents a relatively simple challenge in terms of robot *learning*, but the performance of the learner depends critically on the teacher's skill in understanding what the robot has learned from previous demonstrations.

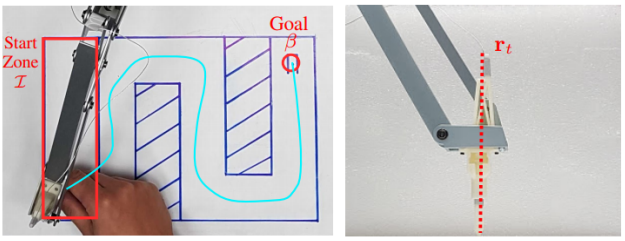
In particular, the efficiency and efficacy of teaching depends on the teacher's ability to ensure the robot *generalises from the teaching examples given*, such that it is able to generate a path from anywhere within the designated start zone to the goal. In other words, $\mathfrak{B}$ represents the space of all such admissible trajectories , and the robot must learn to approximate this from the subset $\mathfrak{b}$ provided by the teacher. To do this, the teacher must form a belief $\tilde{\mathfrak{R}}$ of the robot's actual ability $\mathfrak{R}$ before selecting the demonstration set $\mathfrak{b}$.

It is expected that the teaching quality can be modified and improved by manipulating the teacher's belief space $\mathfrak{M}$, specifically by helping them better understand the learner (influencing $\mathcal{Q}$) to help improve their estimate of the robot's actual ability $\tilde{\mathfrak{R}}$ (2), and the quality of demonstrations (3). Due to the relatively low-dimensional nature of the problem, it is hypothesised that this can be achieved through appropriate design of visual feedback. The following describes the experimental procedure for evaluating this hypothesis and reports results from a group of novice teachers[†].

4.1.1 Hypotheses — The experimental hypotheses are chosen to test whether teacher performance, measured by the metrics (4) and (5), can be improved through appropriate feedback to the teacher. They are formally defined as:

- $\mathbf{H_1}$: *Visualisation of the robot's learning progress results in a significant improvement in teaching efficiency, compared to a no-guidance teaching process.*
- $\mathbf{H_2}$: *A heuristically guided teaching process, which uses visual feedback plus a rule set, results in significantly improved teaching efficiency, compared to a no-guidance teaching process.*

[†]This experiment was approved by a KCL ethics committee, ref. LRS-16/17-3800. Informed consent was obtained from all experimental participants. The data collected for this research is open access, with accreditation, from `http://doi.org/10.18742/RDM01-242`.



(a) (b)

Figure 6. Experimental set up. (a) Participants are asked to guide a lightweight robot through the work space from the start zone to the goal location, avoiding the maze boundaries and the two obstacles (shaded zones). (b) End-effector of the robot indicating the positioning of the TrakStar sensor used for recording data.

$\mathbf{H_3}$: *A heuristically guided teaching process, which uses visual feedback plus a rule set, results in significantly improved teaching efficiency, versus a solely visually guided teaching process.*

$\mathbf{H_1}$ tests whether knowledge of the learner's actual ability $\hat{\mathfrak{R}} \sim \mathfrak{R}$, improves teacher/learner performance beyond the performance achieved when relying on the teacher's understanding $\tilde{\mathfrak{R}}$ alone.

$\mathbf{H_2}$ introduces a heuristic rule set for user guidance, testing whether influencing $\mathcal{Q}$, along with providing knowledge of $\hat{\mathfrak{R}}$ improves teacher/learner performance beyond relying on $\tilde{\mathfrak{R}}$ alone. These rules represent "tips" which an expert might give to a novice. While not always possible, it is reasonably common that guidance can be provided to a learner to avoid undesirable behaviour (such as ambiguous demonstrations), whether that guidance is visual, text or verbal guidance. The distinction here is that the rule based guidance is provided *before* demonstrations are provided, while feedback is provided *after*.

$\mathbf{H_3}$ considers whether modifying $\mathcal{Q}$ along with providing $\hat{\mathfrak{R}}$ improves teacher/learner performance beyond just the feedback of $\hat{\mathfrak{R}}$ alone.

4.1.2 Materials and Methods — Participants in the experiment are asked to teach a robot to navigate its end-effector through a simple two-dimensional maze. To do this, they must provide demonstrations to the robot by gripping its end-effector and manually guiding it through the maze, while avoiding the workspace boundaries and obstacles (see Figure 6a).

More formally, the task skill to be taught to the robot is defined as generating a path which, beginning in a starting area $\mathcal{I}$, passes through an *admissible space* $\boldsymbol{\mathcal{X}}$ to reach a target β, see Figure 6a. $\boldsymbol{\mathcal{X}}$ is defined by a two dimensional bounding rectangle ($20\,cm$ by $30\,cm$), containing the start zone, $\mathcal{I}$ (a $20\,cm$ by $6\,cm$ rectangle) and the target, β (a $0.5cm$ diameter circle). The two obstacles (shown in the figure as shaded blocks) are *not* included in $\boldsymbol{\mathcal{X}}$. Any trajectory $\mathcal{T}_m$ that links points in $\mathcal{I}$ to β, without leaving $\boldsymbol{\mathcal{X}}$, is a member of $\mathfrak{B}$

$$\mathcal{T} \in \mathfrak{B} \quad \text{if} \quad \begin{cases} \text{(i)} & \mathcal{T} \subset \boldsymbol{\mathcal{X}} \\ \text{(ii)} & \mathcal{T}(0) \subset \mathcal{I} \\ \text{(iii)} & \mathcal{T}(T) \subset \beta \end{cases} \tag{11}$$

where $\mathcal{T}_m(0)$ and $\mathcal{T}(T)$ represent the first and last sample in the recorded trajectory respectively. Note that these criteria apply to both demonstrations and task realisations.

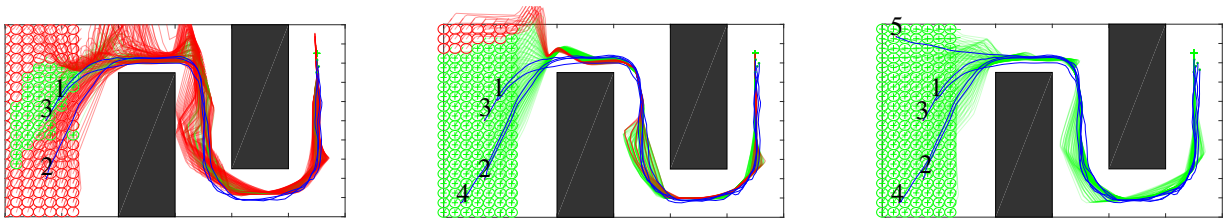
The robot learner used in this experiment is a uFactory uArm Metal, a lightweight robot ($<1\,kg$) with back-driveable motors. During teaching, the robot end effector position is recorded using an NDI TrakStar sensor, (see Figure 6b). This provides $\pm 1.3\,mm$ RMSE positioning accuracy at a sampling rate of $80Hz$.

Using this setup, a data set consisting of M demonstrations is collected during teaching. Each demonstration consists of a trajectory containing T_m samples of the robot state, $\boldsymbol{\xi}_j$, (giving a total of $J = T_m \times M$ samples).

Using this data, the robot uses a Task-Parameterised Gaussian Mixture Model (TP-GMM) (Calinon, 2015) to learn a policy for the demonstrated task. While there are many learning methods which are suitable for this step, TP-GMM is chosen as it has been shown to be particularly effective in generalising from a limited set of demonstrations to unseen conditions. It is important to note that with TP-GMM the robot learner cannot self-refine the learned policy, *i.e.,* the policy is only as good as the data provided by the teacher. Therefore the teacher must make good assessments of the learner's ability, $\tilde{\mathfrak{R}}$, in order to provide suitable set of demonstrations.

In TP-GMM, the task is parameterised by a collection of affine transformations which, in this case, represent a collection of reference frames marking robot end-effector locations and object locations. A local mixture model is learned for the demonstration data in each of these frames of reference. The local mixture models are then combined to the global frame of reference through a product of Gaussians, resulting in a trade-off between the local mixture models which optimises the consistencies observed in data in each frame of reference. Continuous trajectories can then be generated from the global mixture model using Gaussian Mixture Regression (the reader is referred to (Calinon, 2015) for full details of the TP-GMM). In the implementation used here, the mixture models contain $K = 11$ Gaussian components, and the state $\boldsymbol{\xi}_j = (t_j, \mathbf{x}_j^\top) \in \mathbb{R}^{3\times 1}$ is represented with the time t and end-effector position $\mathbf{x}$ for sample j. Four sets of task parameters are used, defining the start, end, and obstacle locations.

Teacher performance is measured according to the efficacy (4) and efficiency (5) of teaching. Here, as $\mathfrak{B}$ and $\mathfrak{R}$ represent continuous spaces in this task, it is necessary to define objective approximations for quantitative comparisons (see §3.1.6). In this experiment, the goal region β is small, so the primary task of teaching is to ensure the learner achieves good generalisation over $\mathcal{I}$. A simple way to approximate $\mathfrak{B}$, therefore, is to discretise $\mathcal{I}$ into a finite set of points, and consider any trajectory that links one of these points to the target β, while meeting the criteria (11), as an element of $\hat{\mathfrak{B}}$. In the results reported here, $\mathcal{I}$ is discretised into a grid of 20×7 points, so $|\hat{\mathfrak{B}}| = 140$. Similarly, $\mathfrak{R}$ is approximated



$m = 3$ (a) $m = 4$ (b) $m = 5$ (c)

Figure 7. Example of a demonstration sequence for $3 \leq m \leq 5$ demonstrations, and the visualisation shown to participants during the visual feedback test conditions. Trajectories in $\hat{\mathfrak{R}}$ are illustrated, with those that meet the criteria (11) coloured green and those that do not coloured red. The teacher's demonstrations (numbered in the order in which they are given) are overlaid in blue.

by the set of trajectories generated by the TP-GMM from each of these points, therefore maximum performance is achieved when $|\hat{\mathfrak{R}} \cap \hat{\mathfrak{B}}| = 140$ giving $\varepsilon = 1$, as shown in Figure 7c.

In two of the experimental conditions, visual feedback is provided to experimental subjects (refer to §4.1.1 and §4.1.3). As a simple means to generate such feedback, the trajectories in $\hat{\mathfrak{R}}$ are presented to the subjects, overlaid onto an image of the maze. These are coloured green if they meet the criteria (11) and red otherwise, as shown in the examples in Figure 7.

4.1.3 Procedure — The following describes the protocol for working with experimental participants.

The experiment is designed as a within-subjects study, so a repeated-measures Analysis of Variance (ANOVA) study is used to determine if any significant effects can be observed in the data. The independent variable in each phase of the experiment is the level of guidance provided to the participant. In all stages of the experiment, the dependent variable is the participant's teaching efficiency, η.

A power analysis for a repeated measures ANOVA, with one group and three measurement levels, indicates that, for a medium effect size (Cohen's $f = 0.3$) and a power of 0.95, the required sample size is $n = 30$ (Faul et al., 2009). In the results reported below, therefore, are for the experiment conducted with 30 participants (17 male, 13 female; ages $\mu = 35.1$, $\sigma = 9.7$). All participants are pre-screened to ensure that they have no background in robotics or machine learning.

The experiment commences with the subject watching an introductory video that (i) explains LfD using a real-world task example, (ii) introduces the objective of teaching the robot to generate trajectories from the start zone to the goal through the maze. Videos are used for all instruction to ensure consistency in the participants' prior knowledge and understanding of teaching phase[‡]. After the introductory video, they are given one minute to familiarise themselves with the robot, where they are free to move it around. After this, the main experiment begins. This is split into three phases, one for each test condition. Each phase is introduced by a video providing instructions, and the participant is given one attempt to provide a set of demonstrations. Participants are not instructed to provide a defined number of demonstrations, as this might provide indirect information about how learning is taking place and prevent natural interaction with the robot.

4.1.4 Conditions — The following are specific details on each test condition used during the experiment.

Condition 1 - No Feedback (NF): In this phase, participants are tested to see if they are able to provide demonstrations with no feedback, *i.e.,* $\mathfrak{r} = \emptyset$, thus relying only on their (uninformed) expectations of the system's behaviour, $\mathcal{Q}$. The instruction video explains the different areas of the task map, provides one basic example of how the task is meant to be performed, and explains that they must provide as many demonstrations as they feel are necessary to enable the robot to perform the task from any point in the start zone.

Condition 2 - Visual Feedback (VF): In this phase, the effect of providing a transparent visualisation of the robot's learning progress is tested. In between each demonstration, the user is provided with the visualisation of learning progress, described in §4.1.2 and shown in Figure 7. The instruction video explains the visualisation, and provides a simple example of what demonstrations look like in the visualisation.

Condition 3 - Visual Feedback, Rule Guidance (VR): In this phase, the participant must additionally follow a set of rules when providing their demonstrations. The rule set is designed to approximately guide users to avoid undemonstrated states and ambiguous demonstrations. The rules are (i) provide one demonstration, starting from anywhere in the start zone, (ii) provide demonstrations within $4\,cm$ of the starting point of the first demonstration, until it is surrounded by successful (*i.e.,* green) test trajectories, and (iii) continue providing demonstrations within $4\,cm$ of the starting point of successful trajectories, in the area with the greatest number of failed (*i.e.,* red) trajectories. The instruction video explains the rules through a single simple example, but avoids dictating exactly how the participant should teach by only showing a small section of the start zone when explaining the rule set.

While random ordering of test conditions is typically employed in within-subjects tests, pilot tests indicated that the learning effect of observing feedback *before* attempting the no feedback condition provided too much information to the participant. Similarly, testing the VR condition before the VF condition often resulted in affecting choice of demonstrations. In the NF case, participants should provide

[‡]The introductory video used in the experiment is provided as a supplementary file to this paper, and is available to view online `https://youtu.be/eafZ1bRAGLM`

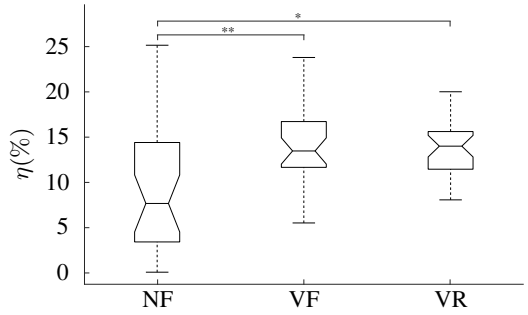


Figure 8. Box plot of results for each test condition. No Feedback (NF), Visual Feedback(VF), Visual Feedback/Rule Guidance (VR). The red lines indicate median values of the teaching efficiency, η, in each case. The bottom and top of the blue boxes indicate the 25^{th} and 75^{th} percentiles, respectively. The upper and lower of the dashed lines indicate maximum and minimum values, respectively.

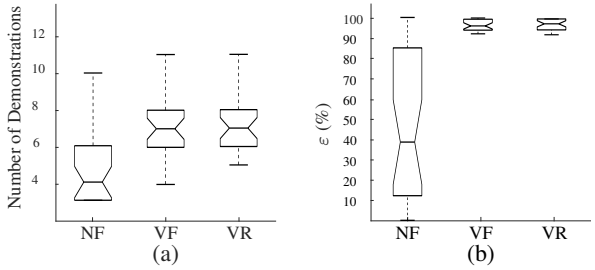


Figure 9. Breakdown of (a) the number of demonstrations provided in each phase, and (b) the percentage of successful trajectories generated, *i.e.,* the *efficacy*, ε.

demonstrations with no knowledge of how learning is taking place. In the VF case, participants should provide demonstrations with no expectation of what qualifies as "good" demonstrations.

4.1.5 Results — Prior to performing a repeated measures ANOVA analysis, the relevant statistical assumptions were all checked and verified for the collected data; continuous dependent variable, at least two groups of independent variables, absenece of outliers, normally distributed dependent variable, and sphericity of data.

The number of demonstrations used for evaluating user performance was taken to be the number of demonstrations required to achieve at least 90% coverage of the test grid, or the maximum number of demonstrations if 90% coverage was not achieved. Analysing the teaching efficiency score, the data indicates a significant effect of the feedback method on the teaching efficiency, $F(2, 58) = 7.952, p = 0.001$.

As a significant effect was observed, a multiple comparisons of means was performed. A significant difference was found between the NF and VF conditions, $p = 0.006$, as well as between the NF and VR conditions, $p = 0.017$. No significant difference was observed between the VF and VR conditions, $p = 0.801$. The median teaching efficiencies, η, are shown in Figure 8. The difference in teaching efficiency medians between the VF (13.4%) condition and the NF (7.6%) condition shows a relative improvement of approximately 180%. A similar teaching efficiency improvement can be observed between the VR (14%) condition and NF.

(a)

(b)

Figure 10. Experimental setup. (a) shows robot is shown in starting position. (b) shows the fiducial markers used to localise the plant tray and disposal bin (green bowl). Red circles in the lower image indicate the *generalisation sampling* test set (see main text).

These results and the statistical analysis show support for $\mathbf{H_1}$ and $\mathbf{H_2}$. This shows a clear benefit in providing the visual feedback to the teacher during LfD, *i.e.,* by helping the teacher to gain an accurate understanding of the learner's current ability, $\tilde{\mathfrak{R}}$, teaching is improved. There is no support found for $\mathbf{H_3}$. As seen in Figure 8, the teaching efficiency of VF and VR are very similar, though it might be noted that the standard deviation of VR is reduced compared to VF, indicating the rule set did offer some assistance to the teacher. This said, the effect of providing the ruleset for the purpose of modfying the teacher's understanding of the learner, captured in $\mathcal{Q}$, is not so apparent.

These findings show how the proposed framework can be used to design an evaluation scheme for the teacher in a given task, as well as tools which are designed to support the teacher's mental representations of the learner to improve teaching performance. See §5 for further discussion.

4.2 Pick-and-Place Experiment

The aim of the second experiment is to evaluate the proposed framework in a real-world scenario, where there are a large number of possible task conditions that may be

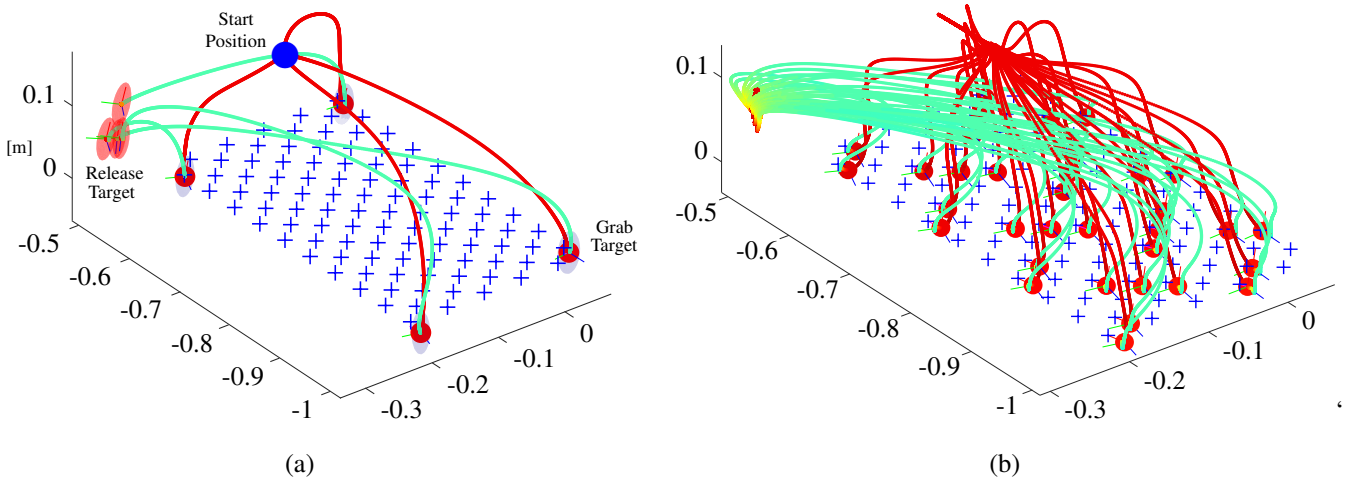


Figure 11. The user provides a set of demonstrations (a), which are used to build a policy that is then used to generate trajectories (b) for all of the targets (blue crosses). The colour of the trajectory indicates the status of the gripper, with red indicating open and green indicating closed. A subset of the generated trajectories is shown in (b) for clarity.

encountered (*i.e.,* large $\mathfrak{B}$ and $\mathfrak{R}$). Typically, this means simple visualisation tools (such as that described in §4.1) are no longer feasible, and alternative approaches must be used. An alternative to visualising task execution in such situations is to have the teacher watch the robot *physically perform the skill* (*i.e.,* using a task-realisation set $\mathfrak{r}$). For the teacher, this presents an additional challenge as they must now rely on a very small set of realisations $\mathfrak{r}$ to form their estimates of $\mathfrak{R}$. In such a scenario, it is unclear (i) how selection of $\mathfrak{r}$ affects novice teachers' understanding of learner abilities $\tilde{\mathfrak{R}}$, and (ii) whether they may benefit from system-selected (as opposed to self-selected) $\mathfrak{r}$.

This experiment explores these questions through the task of plant tray sorting (see Figure 1), whereby the robot must be taught to reach from a start location to a target plant, grasp it, and remove it to a bin (see Figure 10). Note that the high number of degrees of freedom of the robot, and the number of target plants involved make $\mathfrak{B}$ and $\mathfrak{R}$ non-trivial. Additionally, as one realisation (*i.e.,* the picking and removal of a plant from a specific location) takes approximately 25 seconds in the system considered, and there are 100 plants in the tray, feedback to the teacher can be extremely costly in time. This makes the issue of teaching efficiency particularly important for practical deployment of such systems.

4.2.1 Hypotheses — The experimental hypotheses are chosen to test whether teacher skill, as measured by the metrics (4) and (5), are influenced by the choice of feedback in form of the task realisation set $\mathfrak{r}$, and whether the teachers themselves are capable of selecting informative realisations. In forming the set $\mathfrak{r}$, there is flexibility in (i) the order in which demonstrations and realisations are interleaved, (ii) the specific choice of realisations in $\mathfrak{R}$ selected to form $\mathfrak{r}$, and (iii) the responsibility of the teacher to self-select $\mathfrak{r}$, or otherwise.

To examine these issues, the following hypotheses are defined:

$\mathbf{H_4}$: *Feedback through sampling the learnt policy improves teaching efficiency, compared to giving no feedback to the teacher.*

$\mathbf{H_5}$: *Generalisation sampling* (see below) *of the learnt policy improves teaching efficiency, compared to task realisations selected by the teacher.*

$\mathbf{H_6}$: *Generalisation sampling of the learnt policy improves teaching efficiency, compared to the learner simply repeating what it was last shown.*

$\mathbf{H_7}$: *Teaching efficiency improves between episodes as the participant gains understanding of the learning process.*

In this experiment, for $\mathbf{H_5}$ and $\mathbf{H_6}$, the so-called *generalisation sampling* set consists of task realisations for picking the plants located in the four corners of the tray, and one plant located in its centre, see Figure 10. Sampling the learner's skill in these locations gives an estimate of the generalisation capabilities of the robot, since they test its behaviour across the whole space of possible target plants.

$\mathbf{H_5}$, therefore, tests a teacher's ability to interpret feedback which has been designed to test the learner effectively, *i.e.,* $\mathfrak{r} \sim \hat{\mathfrak{R}}$.

$\mathbf{H_6}$ considers the effect of feedback which simply executes the task in the same conditions as shown by the teacher, *i.e.,* $\mathfrak{r} = \mathfrak{b}$.

$\mathbf{H_4}$ considers whether the teacher can effectively teach the learner with no feedback, $\mathfrak{r} = \emptyset$.

$\mathbf{H_7}$ considers whether the feedback acts as a training mechanism in any of the conditions, *i.e.,* whether it modifies $\mathcal{Q}$ in each interaction.

4.2.2 Materials and Methods — Participants in this experiment are asked to teach a robot to grasp plants from a plant tray and dispose of them in a disposal bin (see Figure 10).

More specifically, the skill to be taught is to generate a path which, beginning from a fixed start position, must move along an obstacle-free path toward a specific target plant, into a suitable grasping position (identified by a grasping action being within a threshold distance from the plant). Once there, the robot must grab the plant and move along an obstacle-free path to a bin by the tray. Once there, it must release the plant into the bin (identified as release occurring within a threshold distance from the bin). At all stages of the movement, the gripper must remain in an admissable

space $\boldsymbol{\mathcal{X}}$, which excludes self-collisions and collisions with the table. Examples of good demonstration trajectories are shown in Figure 11a.

These requirements can be formally summarised as

$$\mathcal{T} \in \mathfrak{B} \text{ if } \begin{cases} (i) & \mathcal{T} \subset \boldsymbol{\mathcal{X}} \\ (ii) & d(\mathcal{T}(a_g), \mathbf{b}_g) \leq \delta_g \ , \\ (iii) & d(\mathcal{T}(a_r), \mathbf{b}_r) \leq \delta_r \end{cases} \tag{12}$$

where $d(\cdot, \cdot)$ denotes the Euclidean distance between two points, $\mathcal{T}_m(a_g)$ ($\mathcal{T}_m(a_r)$) is taken as the location of robot end-effector at the grabbing (releasing, respectively) action step, $\mathbf{b}_g$ ($\mathbf{b}_r$) are the grabbing (releasing) target locations, and δ_g (δ_r) is the grabbing (releasing) threshold. The latter thresholds are defined as the mean grab distance observed in the demonstration set, ± 2 standard deviations plus a $1mm$ regularisation term.

In this experiment, the start and end points of the task are fixed positions, and so the teaching must focus on achieving *generalisation over the plant positions in the tray*, which shall form the test set. The test set is again constrained to a finite size, naturally discretised into a set of 100 in the tray grid. Therefore, for this experiment $|\mathfrak{B}| = 100$ and $\max |\mathfrak{R}| = 100$.

As the robot learner, a Rethink Robotics Sawyer robot arm is used, equipped with an Active Robots AR10 hand. The robot arm has 7 degrees of freedom and allows for kinesthetic teaching through gravity-compensated control. The hand has a further 10 degrees of freedom, however, only four of these are used for the grab/release actions. The latter are implemented as a pincer movement (either from the open to closed position, or vice versa), triggered on-demand by the user through press of a button. Detection of the location the plant tray and disposal bin is achieved through fiducial markers (Niekum, 2012), using the in-built cameras of the robot.

During demonstrations, the teacher guides the robot through the required motions for the task using kinesthetic teaching, and the end-effector positions and orientations are recorded through the joint encoders using forward kinematics, with a repeatability of $\pm$1mm. The teacher issues gripper control signals (open/close) using a button on the robot's end-effector cuff, and these are also recorded.

For learning, a policy is learned using TP-GMM with the recorded task demonstrations (see §4.1). The state representation consists of $\boldsymbol{\xi}_j = (t_j, \mathbf{x}_j^p, \mathbf{x}_j^q, x_j^h)^\top$, where t_j is the time stamp, $\mathbf{x}_j^p$ is the position of the robot's end-effector, $\mathbf{x}_j^q$ its orientation and x_j^h is the hand state (either open or closed). The policy is parametrised with three frames of reference: a start frame indicating the initial position of the end-effector, a target frame indicating the position of the target plant, and a goal frame indicating the position of the disposal bin. It uses a mixture of $K = 7$ components, selected based on empirical testing. Examples of task realisations generated by this policy can be seen in Figure 11b.

4.2.3 Procedure — The following describes the protocol for working with the experimental participants.

Participants are screened to ensure they have no prior background in robotics or machine learning. After giving informed consent, they are assigned to one of the four interaction conditions and shown video instructions on their teaching task. Each group begins with the same introduction video, explaining that their goal is to teach the robot to perform a pick-and-place task, and provides a basic explanation of LfD. Following this, to control for any effects relating to the participant's physical coordination skills when interacting with the robot, the participants are allowed a 5 minute familiarisation period. During this time, they have a print-out guidance sheet showing what the buttons on the robot do[§], and are told to try pick-up as many plants from the tray as they can in the time given.

After the familiarisation period, participants are shown a video which explains the teaching procedure and that they will be asked to teach the robot the same task three times. In each teaching session, the participant may provide as many demonstrations as they wish, but are limited to 15 minutes per session. Participants are then shown condition-specific instructions.

In all conditions, prior to providing a demonstration the grab target must be set for the learner. The participant first points to the plant they wish to interact with and the experimenter specifies the target to the robot accordingly on a privately viewed interface (this is done to remove any possibility of the participant being affected by the system software interface). The participant then provides a demonstration of removal of the target plant into the disposal bin. This continues until they decide they have provided enough demonstrations, or the time limit for the session expires. After this point, the robot resets for the next teaching session.

The resultant data is analysed according to a mixed-factors ANOVA. A power analysis indicates that, for a medium effect size (Cohen's $f = 0.3$) and a type I error rate $\alpha = 0.05$, the required sample size is $n = 36$, or 9 per condition (Faul et al., 2009). Accordingly, the results reported below are for a population of 36 participants (18 male, 18 female; ages $\mu = 36.7, \sigma = 10.2$). The latter were recruited from a horticultural production site (domain experts in the task of plant sorting) with all experiments conducted on-site and in a private area[¶].

4.2.4 Conditions — The following describes specific details of each test condition used in the experiment.

Condition 1 - No Feedback (NF) In this condition, participants are only able to observe the robot's task performance after they have provided a complete set of demonstrations (*i.e.,* at the end of a teaching session). No feedback is given during teaching (*i.e.,* $\mathfrak{r} = \emptyset$), representing a situation where the user must rely on their own understanding of the learner's ability, $\tilde{\mathfrak{R}}$, developed exclusively through their knowledge of the demonstrations they provided, $\mathfrak{b}$, and their understanding of how learning is taking place, captured in $\mathcal{Q}$. After a teaching session, they are permitted to view task realisations for self-selected target

[§] See supplementary material *"Robot Controls Guidance Sheet"*.

[¶] This experiment was conducted with ethical approval granted by KCL REC Committee under LRS-17/18-5549. The data collected for this research is open access, with accreditation, from https://doi.org/10.6084/m9.figshare.8953124

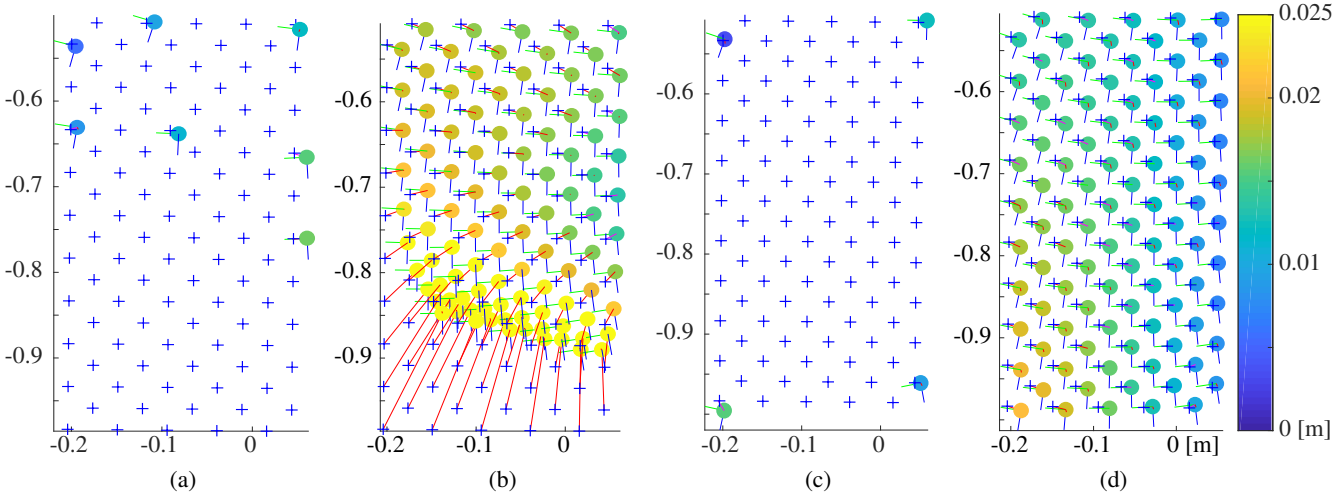


Figure 12. Example of a demonstration and reproduction sets for a case of (a) a poor demonstration set and the resulting model-generated trajectories (b). (c) then shows a good demonstration set and the resulting policy-generated trajectories (d). The blue crosses indicate the tray targets, the red lines indicate the intended target for a grab action, and the shaded circles indicate the location of the grab action for the given target, with the colour representing the distance of the robot grasp from the target at the time of the grab action.

plants. This condition represents a typical naïve approach to LfD, where a person provides some demonstrations of a task, and then observes the robot performing the task.

Condition 2 - Replay Feedback (RF) In this condition, participants are shown a task realisation corresponding to the last demonstration, immediately after it is shown (*i.e.,* during teaching, after each demonstration). Here, the participant forms $\tilde{\mathfrak{R}}$ through observation of $\mathfrak{r}$ and $\mathfrak{b}$, however, as $\mathfrak{r} = \mathfrak{b}$, the feedback gives no indication of generalisation, which the participant must estimate for themselves.

Condition 3 - Batch feedback (BF) In this condition, task realisations for a set of pre-selected test points (*i.e.,* the generalisation sampling set) are shown to the participant after every demonstration. As noted in §4.2.1, these points are selected to give a good approximation of the robot's current ability $\hat{\mathfrak{R}} \sim \mathfrak{R}$. If interpreted by the participant correctly, this feedback gives information as to the extent to which generalisation has occured when forming $\tilde{\mathfrak{R}}$.

Condition 4 - Selected Feedback (SF) In this condition, participants are free to choose when the robot should provide a task realisation, and under what configuration (*i.e.,* which plant location to test) during teaching. Here, the participant forms $\tilde{\mathfrak{R}}$ through their own, self-selected $\mathfrak{r}$.

4.2.5 Results — The data were checked to be compatible with the relevant statistical assumptions. Data in NF, RF, and BF was found to be non-normal using an Anderson-Darling test. ANOVA tests are noted as being robust to violations of normality (Schmider et al., 2010), and considering the excess Kurtosis for the four groups is 1.3695, 1.6042, -0.6387, and 0.9762, respectively (where a normal distribution would have an excess kurtosis value of zero), the violation is considered minor and the data are assumed to follow a normal distribution. Using Mauchly's test for sphericity on the repeated measures model gives $\chi^2(2) = 1.2423, p = 0.5372$, indicating the sphericity assumption is not violated and no data correction is required.

Considering the teaching efficiency for the participant groups, there was a significant main effect observed for the between-subjects factor, indicating a difference was

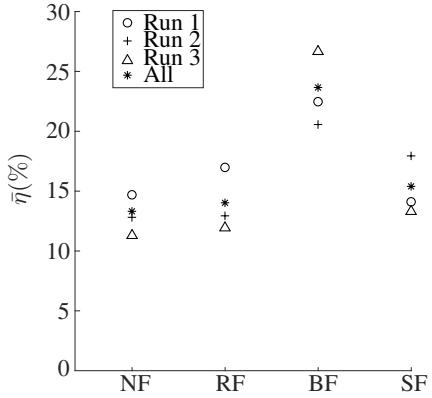
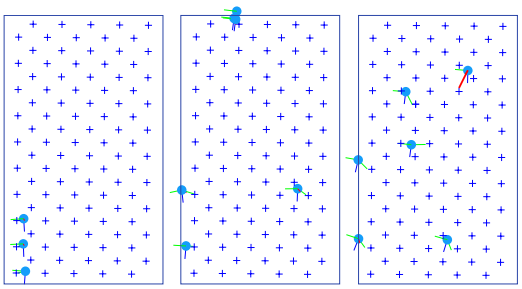


Figure 13. Teaching Efficiency means ($\bar{\eta}$) for inidividual runs as well as the overall mean for the four experimental conditions.

observed between the test conditions (NF, RF, *etc.*), $F(3, 32) = 13.864, p = 5.797 \times 10^{-6}$. As significance was found in the main effect, a multiple comparison of means was performed. This shows that BF provides a 10.76% improvement in efficiency compared to NF ($p = 1.017 \times 10^{-5}$), 9.64% compared to RF ($p = 6.2947 \times 10^{-5}$), and 8.20% compared to SF ($p = 5.8708 \times 10^{-4}$), see Figure 13.

There was no significance found in the within-subjects factor analysis, *i.e.,* there was no support for a learning effect improving performance over each consecutive run over all conditions. However, an interaction effect was observed for the between and within factors, $F(6, 64) = 2.7845, p = 0.01808$).

As an interaction effect was found, a multiple comparison of means conditioned on the within factor is used to consider the *simple main effects*. This multiple comparison indicates that for the first run, no significance in improvement is observed between BF and RF, and no significance is observed between BF and SF. On the third run, significance is observed between BF (27.04%) and each of the other conditions. In this third run, with NF (11.35%), RF (11.94%), and SF (13.7%), the performance of BF represents an average improvement of $\sim 220\%$ relative to each



(a) (b) (c)

Figure 14. Teaching failure examples. (a) undemonstrated states (demonstrations do not cover the space), (b) ambiguous demonstrations (multiple demonstrations co-located at top of grid), (c) errors made during demonstration (demonstration given for wrong target: the red line links the desired target with the demonstration actually given).

condition. On average, across all runs, the mean teaching efficiency of BF represents a $\sim 169\%$ relative improvement over the other conditions.

To gain further insight into the participants' teaching behaviour, the occurance of (i) demonstration errors, (ii) ambiguous demonstrations, and (iii) undemonstrated states are recorded in Table 1. The demonstration error values in Table 1 are gathered by visual inspection of the demonstration sets (see Figure 14c), while the values for undemonstrated states and ambiguous demonstrations are generated by post-processing the data for each participant iteratively to see the performance of the learnt policy after each demonstration is provided, and then applying the definitions of ambiguous demonstrations (9) and undemonstrated states data sets (8). Typical examples of these failures are shown in Figure 14.

Looking at the mean performance levels achieved across all runs for each condition, it can be seen that BF achieved the best overall performance, with 23.54% teaching efficiency. This represents an average improvement of $\sim 169\%$ relative to NF (12.78%), NF (13.89%), and SF (15.34%). The maximum average teaching performance can be found on the third run, where BF achieved 27.04%, representing an average improvement of $\sim 220\%$, relative to NF (11.35%), RF (11.94%), and SF (13.7%).

From these results and the statistical analysis, there is support for $\mathbf{H_5}$, $\mathbf{H_6}$, as BF outperformed the other two feedback methods. As the feedback methods RF and SF did not offer improvement over the no feedback case, it cannot be said that any feedback is better than no feedback and so no support is found for $\mathbf{H_4}$. Finally, while improvement is observed in some cases across subsequent tests, there is no significant evidence to support learning is taking place and thus no support is found for $\mathbf{H_7}$. See §5 for further discussion.

Table 1. Demonstration errors occur when participants provided an example for a plant different to the one they selected. Undemonstrated state counts are calculated using the cardinality of the set defined by (8), and ambiguous demonstrations are found by considering the learner's efficiacy as defined by (9).

	NF	RF	BF	SF
Demonstration Errors	4	0	3	3
Undemonstrated States	499	619	827	485
Ambiguous Demonstrations	54	51	6	39

5 Discussion

The results from both experiments show strong evidence that feedback systems can significantly improve participants' teaching, resulting in improved learner performance on the target tasks. By using the proposed framework to design evaluation and feedback focused on their needs (their interpretation of $\mathfrak{r}$, $\tilde{\mathfrak{R}}$, *etc.*), they develop an understanding of what demonstrations the learner *requires* to learn, without necessarily understanding *how* learning is taking place.

In addition to the basic result of improving teaching performance, a number of insights can be gained from the experimental results. In the first experiment, it can be seen that unguided participants (*i.e.,* those in the NF condition) tend to underestimate the number of demonstrations required to teach effectively, compared to the other conditions (see Figure 9a). This is an indicator that the participants' understanding of how learning is taking place did not match reality, resulting in a poor estimate of the learner's ability ($\tilde{\mathfrak{R}} \neq \hat{\mathfrak{R}}$). Looking at the efficacy (see Figure 9b) it is seen that feedback in the VF and VR cases not only improves performance (higher mean efficacy), but leads to greater consistency (lower standard deviation) among participants, compared to NF. This highlights the difficulty unguided teachers had in providing adequate demonstrations. While some (7 participants) did manage to achieve reasonably good performance in the NF condition, there was wide variation, and the majority performed badly. Conversely, in VF and VR, *all* participants reached close to the maximum efficacy, indicating that the visualisation helped shape their understanding of the teaching task ($\tilde{\mathfrak{B}}$ and $\tilde{\mathfrak{R}}$), and thereby enabled them to provide better demonstrations.

The second experiment also supports these findings, and gives further insight into (i) what kinds of feedback are beneficial, as not all feedback results in improvement, and (ii) the common pitfalls novice teachers encounter. Looking at Figure 13, it can be seen that there is a significant boost to performance observed in the BF condition, but feedback in the RF and SF conditions, offer little improvement over the NF case. The results indicate that for feedback to be of benefit to the teacher, it must offer insight on *how well the learner generalises*, *i.e.,* simply showing the teacher replays of demonstration conditions, $\mathfrak{r} = \mathfrak{b}$ as in the RF case, is not sufficient guidance for teaching. In addition, novice users cannot be expected to know how to test for generalisation without appropriate training, as indicated in the SF condition.

To gain insights into the reasons for this, it is useful to look at participant teaching behaviours. Looking at Table 1, BF featured the fewest ambiguous demonstrations, but also

the most undemonstrated states on average. This suggests that the demonstrations provided by participants under the BF condition are nearly always useful for the learning system, however, the participants are not providing complete demonstration sets. This may either be due to the participants deciding the system has been shown enough information on the task and stopping prematurely (poor $\tilde{\mathfrak{R}}$), or they are not able to provide enough demonstrations in the 15 minute period permitted (feedback in this condition is the most costly in time, with 5 task realisations shown to the teacher after every demonstration). Nevertheless, performance is highest in this condition, suggesting the teaching to be of higher quality.

In addition to these observations, while RF did not provide good overall performance, some benefits to this condition can be seen. Looking at Table 1, participants made no demonstration errors in this condition. Having the robot replay what the participants had just demonstrated appears to help them spot errors as they happen (*e.g.,* providing a demonstration for the wrong target). This suggests that it may be beneficial to design elements of the RF and BF strategies that combine feedback both on immediate errors and learner generalisation.

6 Limitations and Future Work

In both experiments exploring the framework, several task specific choices were made in order to tractably analyse and evaluate the teaching and learning performance. General task performance criteria for assessing the teacher and the learner would prove valuable for deploying LfD more generally.

The framework provides definitions for three common teaching mistakes in LfD, ambiguous demonstrations, undemonstrated states, and incorrect demonstrations. Experimentally, it was observed that incorrect demonstrations had a detrimental effect on learner performance, while the effect of ambiguous demonstrations and undemonstrated states on the learner was less apparent. Future work might consider investigating the sensitivity of learners to these different failure modes.

7 Summary

This paper presents an extended model for LfD, which incorporates the teacher's understanding of, and influence on, the learner. The proposed framework introduces a new space—the teacher's belief space—to the standard view on LfD, highlighting the teacher's understanding of the learner's ability, and how this forms the basis of their interaction. The ability to formalise this relationship, and develop quantitative metrics for its study, is crucial given that learner performance strongly depends on the quality of data provided by the teacher.

The two experiments reported here show the benefit of approaching LfD problems using the proposed framework, by enabling measures for assessing teaching quality, and identifying teaching failures, to be defined. These, in turn, can be used in creating feedback tools to directly influence novice teacher behaviour and guide them toward better teaching practice. Results from the experiments show that, without this guidance, novice teachers struggle to efficiently provide demonstrations which avoid issues like undemonstrated states and ambiguity, even for relatively simple teaching tasks. In contrast, using feedback designed from the perspective of improving teaching quality, as guided by the proposed model, can result in a teaching efficiency improvement of $\sim 169 - 180\%$. These results point to the practical benefits of the proposed model, and it is hoped that this approach to incorporating teachers' thought processes more directly into LfD will help improve novice interactions with LfD systems.

A limitation of the presented framework is the difficulty in identifying the underlying objective task which the user wishes to achieve. Future work on modelling LfD with consideration of the teacher would therefore include overcoming this to autonomously identify optimal test sets for general LfD tasks which are economic with the teaching effort required.

Overall, it is hoped that the presented framework and supporting results highlight the benefit of directly modelling the teacher-robot interactions during LfD, and the resulting new opportunities for evaluating and improving teaching and learning of robotic grasping tasks. By providing a useful structure to LfD problems, feedback tools can be designed to enable novice users to better leverage sophisticated policy learning methods to provide robots with advanced manipulation skills that would be otherwise difficult to train, and difficult to verify learning success.

8 Acknowledgements

This research was supported by the UK Agriculture and Horticulture Development Board (AHDB), under project HNS/PO 194 - *GROWBOT: A Grower-Reprogrammable Robot for Ornamental Plant Production Tasks*, and by the Engineering and Physical Sciences Research Council (EPSRC), under project EP/P010202/ - *SoftSkills: Soft Robotic Skill Learning from Human Demonstration*.

References

Abbeel P and Ng AY (2004) Apprenticeship learning via inverse reinforcement learning. In: *Twenty-first international conference on Machine learning (ICML) '04*. New York, New York, USA: ACM Press, p. 1.

Ahmadzadeh SR and Chernova S (2018) Trajectory-Based Skill Learning Using Generalized Cylinders. *Frontiers in Robotics and AI* 5: 132.

Ahmadzadeh SR, Paikan A, Mastrogiovanni F, Natale L, Kormushev P and Caldwell DG (2015) Learning symbolic representations of actions from human demonstrations. In: *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 3801–3808.

Argall B, Browning B and Veloso M (2007) Learning by demonstration with critique from a human teacher. In: *Proceeding of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. New York, New York, USA: ACM Press, p. 57.

Argall BD, Chernova S, Veloso M and Browning B (2009) A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57(5): 469–483.

Asfour T, Gyarfas F, Azad P and Dillmann R (2006) Imitation Learning of Dual-Arm Manipulation Tasks in Humanoid Robots. In: *6th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. IEEE, pp. 40–47.

Balbach FJ and Zeugmann T (2009) Recent Developments in Algorithmic Teaching. In: *Language and Automata Theory and Applications (LATA)*. Springer Berlin Heidelberg, pp. 1–18.

Billard A, Calinon S, Dillmann R and Schaal S (2008) Robot Programming by Demonstration. In: *Springer Handbook of Robotics*. Berlin, Heidelberg: Springer, pp. 1371–1394.

Billing EA and Hellström T (2010) A Formalism for Learning from Demonstration. *Paladyn, Journal of Behavioral Robotics* 1(1): 1–13.

Cakmak M and Thomaz AL (2011) Active Learning with Mixed Query Types in Learning from Demonstration. In: *Proceedings of the ICML Workshop on New Developments in Imitation Learning*.

Cakmak M and Thomaz AL (2014) Eliciting good teaching from humans for machine learners. *Artificial Intelligence* 217: 198–215.

Calinon S (2015) A tutorial on task-parameterized movement learning and retrieval. *Intelligent Service Robotics* 9(1): 1–29.

Calinon S and Billard A (2007a) Incremental learning of gestures by imitation in a humanoid robot. In: *Proceeding of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. New York, New York, USA: ACM Press, p. 255.

Calinon S and Billard A (2007b) What is the Teacher's Role in Robot Programming by Demonstration? - Toward Benchmarks for Improved Learning. *Interaction Studies. Special Issue on Psychological Benchmarks in Human-Robot Interaction* 8(3).

Calinon S and Lee D (2017) Learning Control. In: *Humanoid Robotics: A Reference*. Dordrecht: Springer Netherlands, pp. 1–52.

Cederborg T, Ming Li M, Baranes A and Oudeyer PY (2010) Incremental local online Gaussian Mixture Regression for imitation learning of multiple tasks. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 267–274.

Cederborg T and Oudeyer PY (2014) A social learning formalism for learners trying to figure out what a teacher wants them to do. *Paladyn, Journal of Behavioral Robotics* 9.

Chernova S and Thomaz AL (2014) *Robot Learning from Human Teachers*. Morgan & Claypool Publishers.

Cho S and Jo S (2013) Incremental Online Learning of Robot Behaviors From Selected Multiple Kinesthetic Teaching Trials. *IEEE Transactions on Systems, Man, and Cybernetics: Systems (SMC)* 43(3): 730–740.

Faul F, Erdfelder E, Buchner A and Lang A (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41: 1149–1160.

Goldman S and Kearns M (1995) On the Complexity of Teaching. *Journal of Computer and System Sciences* 50(1): 20–31.

Goodrich MA and Schultz AC (2007) Human-Robot Interaction: A Survey. *Foundations and Trends® in Human-Computer Interaction* 1(3): 203–275.

Groth C and Henrich D (2014) One-shot robot programming by demonstration using an online oriented particles simulation. In: *IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, pp. 154–160.

Hellström T and Bensch S (2018) Understandable robots - What, Why, and How. *Paladyn, Journal of Behavioral Robotics* 9(1): 110–123.

Hiatt LM, Narber C, Bekele E, Khemlani SS and Trafton JG (2017) Human modeling for humanrobot collaboration. *The International Journal of Robotics Research* 36(5-7): 580–596.

Hoyos J, Prieto F, Alenyà G, Torras C, Prieto F, Alenyà G, Torras C and Torras C (2016) Incremental Learning of Skills in a Task-Parameterized Gaussian Mixture Model. *Journal of Intelligent Robot Systems* 82: 81–99.

Huang Y, Silverio J, Rozo L and Caldwell DG (2018) Generalized Task-Parameterized Skill Learning. In: *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 1–5.

Khan F, Mutlu B and Zhu X (2011) How Do Humans Teach: On Curriculum Learning and Teaching Dimension. In: *Neural Information Processing Systems 24 (NeurIPS)*. pp. 1449–1457.

Kormushev P, Calinon S and Caldwell D (2013) Reinforcement Learning in Robotics: Applications and Real-World Challenges. *Robotics* 2(3): 122–148.

Lewis M, Sycara K and Walker P (2018) The Role of Trust in Human-Robot Interaction. Springer, Cham, pp. 135–159.

Maeda G, Ewerton M, Osa T, Busch B and Peters J (2017) Active Incremental Learning of Robot Movement Primitives. In: *Conference on Robot Learning (CoRL)*. PMLR.

Ng AY and Russell SJ (2000) Algorithms for Inverse Reinforcement Learning. In: *Proceedings of the Seventeenth International Conference on Machine Learning (ICML)*. Morgan Kaufmann Publishers Inc., pp. 663–670.

Nicolescu MN and Mataric MJ (2003) Natural methods for robot task learning. In: *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. New York, New York, USA: ACM Press, p. 241.

Niekum S (2012) ar_track_alvar. URL `http://wiki.ros.org/ar_track_alvar`.

Nikolaidis S, Zhu YX, Hsu D and Srinivasa S (2017) Human-Robot Mutual Adaptation in Shared Autonomy. In: *ACM/IEEE International Conference on Human-Robot Interaction - (HRI)*. New York, New York, USA: ACM Press, pp. 294–302.

Orendt EM, Fichtner M and Henrich D (2016) Robot programming by non-experts: Intuitiveness and robustness of One-Shot robot programming. In: *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 192–199.

Pervez A and Lee D (2017) Learning task-parameterized dynamic movement primitives using mixture of GMMs. *Intelligent Service Robotics* : 1–18.

Schaal S (1996) Learning From Demonstration. In: *Neural Information Processing Systems (NeurIPS)*. pp. 1040–1046.

Schaal S (2006) Dynamic Movement Primitives -A Framework for Motor Control in Humans and Humanoid Robotics. In: *Adaptive Motion of Animals and Machines*. Tokyo: Springer-Verlag, pp. 261–280.

Schmider E, Ziegler M, Danay E, Beyer L and Bühner M (2010) Is It Really Robust? *Methodology* 6(4): 147–151.

Sena A, Zhao Y and Howard MJ (2018) Teaching Human Teachers to Teach Robot Learners. In: *IEEE International Conference*

on Robotics and Automation (ICRA). IEEE, pp. 1–7.

Toris R, Suay HB and Chernova S (2012) A practical comparison of three robot learning from demonstration algorithms. In: *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. New York, USA: ACM Press, p. 261.

Tykal M, Montebelli A and Kyrki V (2016) Incrementally assisted kinesthetic teaching for programming by demonstration. In: *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. ISBN 978-1-4673-8370-7, pp. 205–212. DOI:10.1109/HRI.2016.7451753.

Ureche ALP, Umezawa K, Nakamura Y and Billard A (2015) Task Parameterization Using Continuous Constraints Extracted From Human Demonstrations. *IEEE Transactions on Robotics* 31(6): 1458–1471.

Weiss A, Igelsbock J, Calinon S, Billard A and Tscheligi M (2009) Teaching a humanoid: A user study on learning by demonstration with HOAP-3. In: *The 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 147–152.

Yang XJ, Unhelkar VV, Li K and Shah JA (2017) Evaluating Effects of User Experience and System Transparency on Trust in Automation. In: *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. New York, New York, USA: ACM Press, pp. 408–416.

Zang P, Tian R, Thomaz AL and Isbell CL (2010) Batch versus interactive learning by demonstration. In: *IEEE International Conference on Development and Learning (ICDL)*. IEEE, pp. 219–224.

Zhu X (2015) Machine teaching: an inverse problem to machine learning and an approach toward optimal education. In: *Proceedings of the Twenty-Ninth Conference on Artificial Intelligence (AAAI)*. AAAI Press.

Grollman D. H. and Billard, A. G. (2012) Robot Learning from Failed Demonstrations. In: *International Journal of Social Robotics*. pp. 331–342.